Fluctuations in single-shot ϵ -deterministic work extraction

Sina Salek[1,2,*] and Karoline Wiesner[2,3]
[1]*Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong*
[2]*School of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom*
[3]*Bristol Centre for Complexity Sciences, University of Bristol, 1-9 Old Park Hill, Bristol BS2 8BB, United Kingdom*
(Received 14 August 2017; published 13 November 2017)

There has been an increasing interest in the quantification of nearly deterministic work extraction from a finite number of copies of microscopic particles in finite time. This paradigm, the so-called single-shot ϵ-deterministic work extraction, considers processes with small failure probabilities. However, the resulting fluctuations in the extracted work entailed by this failure probability have yet to be studied. In the standard thermodynamics, paradigm fluctuation theorems are powerful tools to study fluctuating quantities. Given that standard fluctuation theorems are inadequate for a single-shot scenario, here we formulate and prove a fluctuation relation specific to the single-shot ϵ-deterministic work extraction to bridge this gap. Our results are general in the sense that we allow the system to be in contact with the heat bath at all times. As a corollary of our theorem, we derive the known bounds on the ϵ-deterministic work.

DOI: 10.1103/PhysRevA.96.052114

I. INTRODUCTION

Thermodynamics has been historically developed as a discipline that deals with a large number of systems, e.g., an ensemble of particles in a box. In recent years, however, there has been significant interest in studying the relationship between thermodynamics and statistical physics, for a finite number of copies of particles and a finite application of operations on them. This limit is called single-shot thermodynamics [1–4]. A particularly interesting question in single-shot thermodynamics is the extraction of work from systems that are out of equilibrium [3,4]. In order to formulate work extraction, we first need to define work. The first law of thermodynamics splits energy into an ordered form and a disordered form. The ordered form of energy is called work. The challenge faced in the quantification of work in the single-shot regime is that, for small ensembles of microscopic systems, fluctuations dominate. Hence, determining the typical behavior of systems, for which one would need large ensembles, is challenging. Thus, it is not clear how one can quantify ordered energy, i.e., work as perceived in standard thermodynamics. A way around this is to design a process which uses an amount of energy to lift the state of an external system, called the "battery," from a single energy level to another single energy level. This way one can define a notion of work that can be quantified in the single-shot regime, and that also corresponds to the notion of work as understood in standard thermodynamics. Nevertheless, since fully deterministic work extraction can be too stringent a constraint, one might also allow processes that fail with a small probability $\epsilon \ll 1$. Such a process is called "ϵ-deterministic," which is to say, "nearly deterministic." Horodecki and Oppenheim [3] discussed a scenario involving a system out of thermal equilibrium and block diagonal in the energy eigenbasis, coupled to a bath and battery. The task is to lift the state of the battery from the ground state to its excited state ϵ-deterministically. If the task is done successfully, a certain amount of work is stored in the battery. In this scenario of a small failure probability $\epsilon \ll 1$, they find an upper bound for the work extracted. The existence of a failure probability entails some fluctuations in the work extracted. Experiments which fail to extract exactly the desired amount of work are dismissed entirely. What has not been considered so far is that rather than dismissing these experiments, one can quantify the alternative amount of work that has been extracted and how it fluctuates. In this paper, we formulate and prove a theorem (Theorem 1) that characterizes the fluctuations of the extractable work due to the failure probabilities in single-shot thermodynamics.

To answer this, we use an idea developed in another domain of statistical physics. In statistical physics, powerful tools called fluctuation theorems [5] have been developed to characterize fluctuations in quantities such as work or entropy. They put restrictions on the probability density function of these fluctuating quantities. In the case of work, for instance, these theorems compare the probability of work cost of a thermodynamic process with the probability of work gain of its reverse process. An example of fluctuation theorem is the Crooks work relation [6]. Crooks theorem relates the probability $P(w,\mathcal{P})$ of work cost w of driving a thermal state to one that is out of equilibrium, to the probability $P(-w,\mathcal{P}^{\text{rev}})$ that the time reverse of the forward process costs the same amount of work, where $\mathcal{P}$ and $\mathcal{P}^{\text{rev}}$ are the process and its inverse, respectively. The relation is formally expressed as $\frac{P(w,\mathcal{P})}{P(-w,\mathcal{P}^{\text{rev}})} = e^{\beta(w-\Delta F)}$, where β is the inverse temperature of the environment divided by the Boltzmann constant, and ΔF is the equilibrium free-energy difference of the initial and final state. A thermodynamic process, there, is defined by changing the Hamiltonian of the system at an arbitrary speed, according to a specific trajectory of the Hamiltonian. And the reverse process is one where the Hamiltonian is brought back to the initial setting, according to the reversed trajectory of the forward process. The ratio of these two probabilities equals the exponential of the work dissipated in the process, $w_{\text{diss}} = w - \Delta F$. Therefore, in a process where no work is dissipated, e.g., an isothermal expansion, these two probabilities are the same and no fluctuation in work occurs. As we see, this theorem characterizes fluctuations in a thermodynamical quantity that meets the conditions of

*salek.sina@gmail.com

the theorem, by comparing the probability of the forward process to that of the backward process. In this paper, we show that the main ideas of fluctuation relations can be used to characterize the fluctuations arising from the ϵ failure probabilities. However, since the currently existing fluctuation theorems have different assumptions from the ones in the single-shot work extraction, we formulate our theorem specifically for the single-shot scenario. This theorem will put the notion of fluctuations in ϵ-deterministic work extraction, and that in standard thermodynamics on an equal footing.

To formulate a fluctuation theorem for single-shot thermodynamics, we relax the original assumption of the Crooks work relation that the system, both in the forward process and its inverse, starts in a state of thermal equilibrium with its environment. Our setup, in keeping with the setting in Ref. [3], consists of a work system, a battery, and a thermal bath. The work system remains in contact with the bath throughout the process. The protocol extracts work W by performing some operations on the work system and as a result lifts the battery to the excited state $|W\rangle\langle W|$ with $\epsilon \ll 1$ failure probability. As a corollary of our result, we obtain the bounds of the extractable work found by Åberg [4].

It is worth noticing that as a by-product of choosing the present setting, we avoid a potential complication in the fluctuation analysis of open systems. In particular, thermodynamic work W is measured in a number of different ways in the context of fluctuation relations. Predominantly, the so-called scheme of the two-measurement protocol (TMP) is used where the eigenvalues of the initial and final Hamiltonians are measured, and work W is defined as the difference between the two eigenvalues. Considering the first law of thermodynamics, we observe that for closed systems with no heat flow, this quantity is the same as total work. However, in the case of open systems with strong coupling, TMP does not provide a good measure of work. In our setup, all of the work goes to lifting the state of the battery, and, despite the fact that the heat bath is attached to the work system at all times, a simple act of measuring the initial and final eigenvalues of the battery gives a correct measure of the work.

In the next section we give the technical preliminaries before presenting the main theorem (Theorem 1) in Sec. III. The proof of Theorem 1 is given in the Appendix.

II. PRELIMINARIES

The tool with which we choose to prove the theorem is the so-called resource theory of states out of thermal equilibrium. Resource theories [7,8] are mathematical formalisms, designed to answer questions such as "what state can be prepared in a given physical situation?" or "is a transition from one state to another possible, and if yes, what is the transition probability?" By answering these questions, resource theories allow us to study what happens to the state of a system in a physical situation in full generality. Resource theories achieve this by restricting the allowed operations to a predefined set. For a given set of operations one can prepare a set of states for free. Any state outside that set is considered a resource. For instance, it is a well-known fact that if we restrict ourselves to local operations and classical communications (LOCC), we can create separable states. For LOCC, therefore, an entangled state is considered a resource. The resource theory of athermal states is defined by restricting our allowed operations to the so-called thermal operations. Thermal operations, as explained concisely in Ref. [9], are the set of maps characterized by the following rules: (i) A system with any Hamiltonian in the Gibbs state of that Hamiltonian can be added, (ii) any subsystem can be discarded through tracing out, and (iii) any energy-conserving unitary, i.e., those unitaries that commute with the total Hamiltonian, can be applied to the global system. For these operations a state out of equilibrium is a resource and can be utilized to extract work. The commutation condition is imposed to ensure that total energy is conserved. As we see, the resource theory of athermal states provides a framework wherein one can study the most general thermodynamical state transformations that respect energy conservation. Specifically, in our theorem, as we are interested in thermodynamical processes that extract a certain amount of work, we use this framework to identify all possible processes that meet our conditions. In our setting we are interested in making as many abstractions as possible to study the fluctuations induced by the failure probabilities. Therefore, we do not consider generalizations to the cases where other factors such as magnetic fields or chemical potentials are involved.

In our setup, initially the total system consists of the work system in state ρ_s from which we intend to extract work, a bath in state τ_{bath}, and a battery in its pure ground state. The battery is a finite-dimensional many-level system, used for storing the ordered energy, i.e., work. A many-level battery is needed in our case in order to store the different possible amounts of the fluctuating work. A thermal operation is used to extract work from the work state. The thermal operation is generated by a global unitary operator on the initial state of the total system. This operator creates correlations between the work state, bath, and battery. As we can accept a small failure probability, the final state of the system can be slightly different every time we run the protocol. In particular, the battery may be charged to the point W^δ, where δ here signifies different excited levels of the battery. The thermal operation is chosen such that it aims to transform the initial state of the work system to the Gibbs state τ_S. However, again, due to the failure probability, the work system might end in a state that is not exactly Gibbs, and we call this state τ_S^δ. A more detailed study of the final state in the setting we choose can be found in Ref. [2]. To formulate our fluctuation theorem we also need to specify a backward process. The backward process is generated by the unitary operator, the inverse of the global unitary used in the forward process. We choose the initial state of the backward process to be the correlated state of the work system and bath, such that the work state is left in a Gibbs state, after tracing out the state of the bath, as explained below. The initial state of the battery in the backward process is the same as its final state in the forward process. This guarantees that the work extracted in the forward process and the work cost in the backward process are the same. The forward and backward processes are illustrated in Fig. 1.

Our choice of the initial state of the backward process is due to a subtlety inherent in single-shot thermodynamics. In Ref. [8], it was shown that there is an intrinsic irreversibility in the single-shot scenario in the sense that if a resource ρ can be converted to a resource σ, in general, one cannot

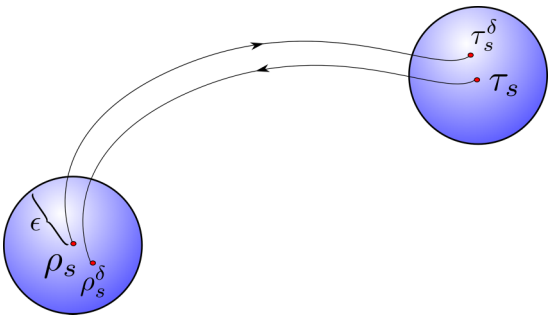

FIG. 1. In the forward process, the system is initially prepared in the state ρ_s. Evolving the system and the bath together according to some global unitary operator correlates the state of the two. The transformation is such that a work W^δ is extracted and the final state of the system τ_s^δ is close to the Gibbs state. The backward process is defined by the dual operator to the unitary in the forward process, and the initial state of the bath and the system are correlated, such that tracing out the state of the bath leaves the system in the Gibbs state. The final state of the backward process is in the ball of ϵ-close states to the initial state ρ_s.

assume that the resource σ can be converted back to the initial resource ρ under the same conditions. In particular, in the first part of Ref. [3], the maximum work extractable in the work extraction setting above was found, where work was extracted by transforming the resource state ρ into the Gibbs state. In the second part, a reverse of that process was defined as a task of starting with a bath, a resource system in Gibbs state, and a battery, all initially in tensor product with each other. It was shown that the work cost of returning the Gibbs state back to ρ is greater than the work extracted in the first part in the single-shot scenario. This is due to the detailed correlations that are created between the system and the bath after the application of the global unitary that generates the thermal operations, which are significant for single-shot work extraction. Therefore, the backward process for a fluctuation relation described above has to begin in a correlated state of the resource system and the bath, such that the correlations are produced by the operations of the kind used in the forward process.

In our theorem we use the ϵ-smooth Renyi relative entropy of order $\alpha = 0$. The Renyi relative entropy of order 0 is defined as $D_0(\rho_s||\tau_s) := -\ln \mathrm{Tr}[\rho_s^0 \tau_s]$, with ρ_s^0 being the support of ρ_s. The ϵ-smooth version of this entropy is defined by maximizing it over the distributions that are ϵ-close to the state of the system, formally defined as

$$D_0^\epsilon(\rho_s||\tau_s) := \max_{\bar{\rho}_s \in \mathcal{B}^\epsilon(\rho_s)} D_0(\bar{\rho}_s||\tau_s), \tag{1}$$

with

$$\mathcal{B}^\epsilon(\rho_s) := \{\bar{\rho}_s \geqslant 0 : ||\bar{\rho}_s - \rho_s||_1 \leqslant \epsilon, \mathrm{Tr}(\bar{\rho}_s) \leqslant \mathrm{Tr}(\rho_s)\} \tag{2}$$

defining the ball of ϵ-close distributions. This entropy was shown to act as a nonequilibrium free-energy difference in the single-shot regime [3,4].

III. THE MAIN THEOREM

We now state the main result, a fluctuation theorem relating the ratio between the probability of extracting work W in an ϵ-deterministic forward process and the probability of putting work W back into the system in a reversed process.

Theorem 1. Consider a total system $\tau_{\mathrm{bath}} \otimes \rho_s \otimes |0\rangle\langle 0|$ consisting of a work system in some state ρ_s, diagonal in energy eigenbasis, a thermal bath at temperature T in state τ_{bath}, and a many-level battery system in its pure ground state $|0\rangle\langle 0|$. Furthermore, assume an ϵ-deterministic forward process $\mathcal{P}^\epsilon$, defined by a global unitary on the total system that commutes with its Hamiltonian, and a reverse process $\mathcal{P}^{\epsilon,\mathrm{rev}}$, defined by inversing the global unitary of the forward process. The initial state of the backward process is a correlated state of the bath and the resource system, such that the work state is in a Gibbs state τ_S after tracing out the bath. With these assumptions, the ratio of work extraction probability $P(W,\mathcal{P}^\epsilon)$ to work cost probability $P(-W,\mathcal{P}^{\epsilon,\mathrm{rev}})$ is given by

$$\frac{P(W,\mathcal{P}^\epsilon)}{P(-W,\mathcal{P}^{\epsilon,\mathrm{rev}})} = e^{\beta W + \ln(1-\delta) - D_0^\epsilon(\rho_s||\tau_s)}, \tag{3}$$

where W is the work extracted. $\beta = 1/kT$, where k is the Boltzmann constant, and T is the temperature. $D_0^\epsilon(\rho_s||\tau_s)$ is the smooth version of Renyi relative entropy of order $\alpha = 0$. $\ln(1-\delta)$ is the amount of deviation from the maximum extractable work, with $0 \leqslant \delta \leqslant \epsilon$.

We provide the proof of Theorem 1 in the Appendix. Theorem 1 characterizes the fluctuations in the extracted work due to the failure probability in the single-shot ϵ-deterministic work extraction scenario. It states that if a quantum resource is used to extract work in an ϵ-deterministic process by thermal operations, the probability of the work extracted is related to the probability of the work put back into the system in its reversed process by an exponential factor. The term in the exponent of Eq. (3) is the analog of dissipated work. This characterizes a type of irreversibility in the finite-run behavior of a microscopic system in the single-shot regime, akin to that of the Crooks relation. As discussed before, the final state of the system in the forward process may be such that the state of the system is not exactly Gibbs. The term $\ln(1-\delta)$ signifies this deviation. We show the details of this in the Appendix.

In the following corollary, we shall observe how a known bound of work extraction follows from our result.

Corollary 1. The work that can be extracted from a system out of equilibrium by thermal operations in an ϵ-deterministic process is bounded by

$$\begin{aligned} kT\, D_0^\epsilon(\rho_s||\tau_s) + kT\ln(1-\epsilon) - kT\ln(1-\delta) \leqslant W \\ \leqslant kT\, D_0^\epsilon(\rho_s||\tau_s) - kT\ln(1-\epsilon) - kT\ln(1-\delta), \end{aligned} \tag{4}$$

with $0 \leqslant \delta \leqslant \epsilon$.

Proof. To see this, we first derive the lower bound and then the upper bound as follows. For the lower bound we multiply both sides of Eq. (3) by $P(-W,\mathcal{P}^{\epsilon,\mathrm{rev}})$. From the fact that $P(W,\mathcal{P}^\epsilon) \geqslant 1-\epsilon$, we have

$$1 - \epsilon \leqslant P(-W,\mathcal{P}^{\epsilon,\mathrm{rev}}) e^{\beta W + \ln(1-\delta) - D_0^\epsilon(\rho_s||\tau_s)} \tag{5}$$

$$\leqslant e^{\beta W + \ln(1-\delta) - D_0^\epsilon(\rho_s||\tau_s)}, \tag{6}$$

where in the second inequality we used the fact that any probability is less than or equal to one. Taking the logarithm of Eq. (6), the lower bound in the statement of Corollary 1 follows.

For the upper bound we rewrite Eq. (3) as

$$\frac{P(-W,\mathcal{P}^{\epsilon,\mathrm{rev}})}{P(W,\mathcal{P}^{\epsilon})} = e^{-\beta W-\ln(1-\delta)+D_0^{\epsilon}(\rho_s||\tau_s)}. \tag{7}$$

Following the same procedure as above, here we have

$$1-\epsilon \leqslant P(-W,\mathcal{P}^{\epsilon})e^{-\beta W-\ln(1-\delta)+D_0^{\epsilon}(\rho_s||\tau_s)} \tag{8}$$

$$\leqslant e^{-\beta W-\ln(1-\delta)+D_0^{\epsilon}(\rho_s||\tau_s)}. \tag{9}$$

Taking the logarithm gives the upper bound of Corollary 1. ■

In Ref. [4], Åberg gives the bounds on work extraction in ϵ-deterministic work extraction as

$$kT\,D_0^{\epsilon}(\rho_s||\tau_s) \leqslant W \leqslant kT\,D_0^{\epsilon}(\rho_s||\tau_s) - kT\ln(1-\epsilon). \tag{10}$$

Notice that the upper bound in Eq. (10) is the special case of Eq. (4) with $\delta = 0$. The lower bound is the special case of Eq. (4) with $\delta = \epsilon$.

IV. DISCUSSION

In the discussion section of Ref. [4], Åberg leaves the question of the link between the fluctuations of the ϵ-deterministic work extraction paradigm to that of the fluctuation theorems paradigm open. Theorem 1 above has answered this question. One can see, as we show in the Appendix, that in the absence of any fluctuations of the former type (the case where ϵ and δ are zero), the dissipated work goes to zero and the forward and backward probabilities take the same value. This means there will be no fluctuations of the latter type. It is only through the introduction of the failure probabilities that we obtain a fluctuation theorem, hence demonstrating the direct connection between the two. We have derived an equality in the same way the Crooks relation does for the work probability density function in the standard account of thermodynamics. In the recognition that the fluctuations induced by the failure probabilities of the ϵ-deterministic work extraction come from a different source than those in nonequilibrium processes that are captured by the Crooks relation, we have made the assumption that the processes we consider are the ones that extract the maximum amount of work permitted by the second law of thermodynamics. In that case, since there is no dissipated work due to the nonequilibrium process, there will be no fluctuations of that type. The relation in Theorem 1 quantifies the ratio of the probability of work extraction in a process to that of work cost in its backward process, in terms of the smooth Renyi relative entropy of order $\alpha = 0$. As a corollary, we find the known bounds of work extraction protocols. The necessity of formulating and proving this variation of fluctuation theorem is that the single-shot regime has different assumptions to the standard thermodynamics, which renders the standard fluctuation relations inadequate for our purposes. In the standard fluctuation relations the work is determined by making several energy measurements at the beginning and the end of the processes and comparing the energy eigenvalues. This requires many measurements. In the single-shot scenario we require to be able to determine work by a single measurement. In our setting this is done by letting the process store all of the work in the battery. Then, the work is measured by a single energy measurement on the battery state. Given the record in the experimental applications of the present fluctuation theorems [10–12], we believe our work will contribute to the experimental results on single-shot thermodynamics by providing a method to measure smooth Renyi entropies. As an outlook on further theoretical extensions of this work, one can think of generalizations or restrictions on the allowed operations, as well as a consideration of quantumness. For instance, Faist *et al.* have shown that Gibbs-preserving maps outperform thermal operations in the quantum regime [13]. A further investigation into the hierarchy of operations and the possible exploitation of quantum correlations and catalysts merits an extended discussion. Furthermore, it would be interesting to apply a version of the asymptotic equipartition theorem [14] to investigate the relationship between our theorem and the ones in the standard statistical mechanics. In this paper, we have focused on the notion of ϵ-deterministic work. However, probabilistic work in the single-shot scenario is also an interesting concept and has been recently studied [9].

Note added. Recently, we became aware of a work by Dahlsten *et al.* [15], who obtained similar results independently, using a different setup and different starting assumptions.

ACKNOWLEDGMENTS

We would like to thank Gonzalo Manzano, Johan Åberg, Jens Eisert, Oscar Dahlsten, Rodrigo Gallego, Henrik Wilming, and Albert Werner for their comments and discussions. Part of this work was supported by the COST Action MP1209 "Thermodynamics in the quantum regime," Hong Kong Research Grant Council through Grant No. 17326616, National Science Foundation of China through Grant No. 11675136, HKU Postdoctoral Fellowship. K.W. acknowledges funding through EPSRC (EP/E501214/1).

APPENDIX: PROOF OF THEOREM 1

Here, we prove Theorem 1 by deriving Eq. (3). However, to begin with, we consider the simple case of a process which extracts the amount of work W from the work system with probability one, i.e., with failure probability $\epsilon = 0$. For such a case, the following lemma holds. This does not give a fluctuation theorem directly, as without a failure probability there is no fluctuation. However, it is an important building block in proving the theorem. In this lemma, we use Renyi relative entropy of order $\alpha = 0$ defined as $D_0(\rho_s||\tau_s) := -\ln \mathrm{Tr}[\rho_s^0\tau_s]$, with ρ_s^0 being the support of the initial state and τ_s the thermal state of the system. In the following lemma, we adopt the same sign convention for work as used in Refs. [3,4]. Here, we have the same assumptions as in Ref. [3] for the allowed operations to be thermal operations, which are obtained by performing some global unitaries on

the total initial state. In the following lemma, the reverse process is obtained from the inverse of the global unitary in the forward process. The initial state of the reverse process, here, is set to be the same as the final state of the forward process. Our assumptions on the bath are also the same as Ref. [3], namely that (i) the spectrum of the heat bath is continuous, i.e., for an energy of the heat bath E_R and two arbitrary energies of the system, E_S and E'_S, there exists E'_R, such that $E_R + E_S = E'_R + E'_S$, and (ii) around the energy E the degeneracies can be written as

$$g(E+\Delta E) = e^{S(E+\Delta E)} = e^{S(E)+\Delta E\frac{\partial S(E)}{\partial E}+O(\Delta E^2)} \quad \text{(A1)}$$

$$= g(E)e^{\beta\Delta E+O(\Delta E^2)}, \quad \text{(A2)}$$

where $S(E) := \ln g(E)$ and $\beta := \frac{\partial S(E)}{\partial E}$. Notice that for a large enough bath the second- and higher-order terms in Eqs. (A1) and (A2) can be neglected, which is the assumption we adopt throughout this paper.

Lemma 1. Consider a total system $\tau_{\text{bath}} \otimes \rho_s \otimes |0\rangle\langle 0|$ consisting of a work system in some arbitrary state ρ_s, diagonal in energy eigenbasis, subject to a time-independent Hamiltonian, a thermal bath at temperature T in state τ_{bath}, and a two-level battery system in its pure ground state $|0\rangle\langle 0|$. Here, the work is extracted by lifting the state of the battery from its ground state to the pure excited state $|W\rangle\langle W|$. The gap between the ground and the excited state of the battery is set to W. Furthermore, assume processes $\mathcal{P}$ and $\mathcal{P}^{\text{rev}}$ consisting of thermal operations only. Then,

$$\frac{P(W,\mathcal{P})}{P(-W,\mathcal{P}^{\text{rev}})} = e^{\beta W - D_0(\rho_s||\tau_s)}, \quad \text{(A3)}$$

with $\beta = 1/kT$, where k is the Boltzmann constant.

Proof. In Ref. [3], it was shown that for a fixed total energy E, if the bath is very large compared to the system, the state $\tau_{\text{bath}} \otimes \rho_s \otimes |0\rangle\langle 0|$ can be written as $\bigoplus_{E_s} \eta^R_{E-E_s} \otimes \Pi_{E_s}\rho_s\Pi_{E_s} \otimes |0\rangle\langle 0|$, where $\eta^R_{E-E_s} := \frac{\mathbb{I}_R^{E-E_s}}{g_R(E-E_s)}$, $g_R(E-E_s)$ is the degeneracy of the bath, and the identity $\mathbb{I}_R^{E-E_s}$ acts on a $g_R(E-E_s)$-dimensional space. Similarly, the final state can be written as $\bigoplus_{E_s^f} \eta^R_{E-E_s^f-W} \otimes \Pi_{E_s}\tau_s\Pi_{E_s} \otimes |W\rangle\langle W|$. Since the total system is block diagonal in the energy eigenbasis, we can treat the eigenvalues of it as probabilities. Denote the energy levels of the initial and final state of the work system by E_i and E_j, respectively. The eigenvalues of the initial and final state are $\frac{P(E_i)}{g(E-E_i)}$ and $\frac{P(E_j)}{g(E-E_j-W)}$, respectively. Denote the transition current, i.e., the number of eigenstates that go from E_i to E_j, by $k_{i\to j}$. Then, occupation probabilities of the final state of the system are given by

$$P(E_j) = \sum_i k_{i\to j}\frac{P(E_i)}{g(E-E_i)}. \quad \text{(A4)}$$

Each summand on the right-hand side (RHS) is the joint probability of the system initially occupying the ith energy level and ending up in the jth energy level in the final state. Since the protocol stores the energy difference between the initial and final state, i.e., the work W, in the battery state the probability of a value of work $P(W,\mathcal{P})$ can be calculated by summing the RHS over j. Notice that the total number of eigenstates in the initial energy level E_i is

$$d_i = \sum_j k_{i\to j} = g(E-E_i), \quad \text{(A5)}$$

and that the total number of eigenstates in the final energy level E_j is

$$d_j = \sum_i k_{i\to j} = g(E-E_j-W). \quad \text{(A6)}$$

We also need to consider the backward process. Recall that the global unitary for the forward process gives transition currents $k_{i\to j}$ as the rate at which energy eigenstates are transformed from E_i to E_j. Therefore, as we are restricting ourselves to the diagonal case, this current can always be used to directly characterize the backward transition currents $k^{\text{rev}}_{j\to i}$ as the rate at which the energy eigenstates are transformed back from E_j to E_i. Notice for the reversed process we have

$$\sum_i k^{\text{rev}}_{j\to i} = g(E-E_j-W), \quad \text{(A7)}$$

which is the number of eigenstates in the final state of the forward process. In the setting of this lemma, this is also the number of eigenstates of the initial state of the backward process.

As in the standard Crooks relation, it is also important to fix the initial state of the backward process. For this lemma we use the final state of the forward process as the initial state of the backward process. Here, the final state of the bath and the work system is a correlated state such that the system is in a thermal state after tracing out the bath.

Remark 1. Formally speaking, a thermal operation T is generated by a global unitary U, acting jointly on the reservoir, system, and battery, such that $\sigma_{sb} := T(\rho_{sb}) = \text{Tr}_R(U\rho_{Rsb}U^\dagger)$, where ρ_{Rsb} is the total state of the reservoir, system, and battery, with a reduced state ρ_{sb}. The restriction on this global unitary is that it has to commute with the total Hamiltonian of the reservoir, system, and battery. In the current case, the global unitary sends the initial state $\tau_{\text{bath}} \otimes \rho_s \otimes |0\rangle\langle 0|$ to the final state,

$$U\tau_{\text{bath}} \otimes \rho_s \otimes |0\rangle\langle 0|U^\dagger := \sigma_{\{\text{bath}\}s} \otimes |W\rangle\langle W|, \quad \text{(A8)}$$

such that the reduced state of the system σ_s is the Gibbs state. These are the same initial and final states considered in the exact transformation case in Ref. [3]. Indeed, in Ref. [3], it was shown that a reverse process starting with a fresh bath will cost more energy than the forward process. However, defining the backward process as the one produced by the inverse of the forward unitary, i.e., $U^\dagger$, and starting from the final state of the forward process, clearly completely reverses the process as $U^\dagger U\tau_{\text{bath}} \otimes \rho_s \otimes |0\rangle\langle 0|U^\dagger U = \tau_{\text{bath}} \otimes \rho_s \otimes |0\rangle\langle 0|$. We shall make a remark on the backward process with nonzero epsilon later in this Appendix, after we provide the proof of the theorem.

Below, we calculate the ratio of the work cost probability to that of work gain probability for such a process,

$$\begin{aligned}\frac{P(W,\mathcal{P})}{P(-W,\mathcal{P}^{\text{rev}})} &= \frac{\sum_{i,j} k_{i\to j}\frac{P(E_i)}{g(E-E_i)}}{\sum_{i,j} k^{\text{rev}}_{j\to i}\frac{P(E_j)}{g(E-E_j-W)}}\\ &= \frac{\sum_i g(E-E_i)\frac{P(E_i)}{g(E-E_i)}}{\sum_j g(E-E_j-W)\frac{P(E_j)}{g(E-E_j-W)}}\\ &= \frac{\sum_i\sum_j k^{\text{rev}}_{j\to i}\frac{P(E_j)}{g(E-E_j-W)}}{\sum_j P(E_j)}. \end{aligned} \tag{A9}$$

Using the fact that the system ends up in the Gibbs state, together with the assumption on the bath in Eq. (A2) that $g(E-E_j)=g(E)e^{-\beta E_j}$, one can rewrite the term in the numerator as follows,

$$\begin{aligned}\sum_i\sum_j k^{\text{rev}}_{j\to i}\frac{P(E_j)}{g(E-E_j-W)} &= \sum_i\sum_j k^{\text{rev}}_{j\to i}\frac{1}{g(E-W)Z}\\ &= \sum_i \frac{g(E-E_i)}{g(E-W)Z}, \end{aligned} \tag{A10}$$

where Z is the partition function of the Gibbs state.

Inserting Eq. (A10) into Eq. (A9) and simplifying the fraction, we find

$$\frac{P(W,\mathcal{P})}{P(-W,\mathcal{P}^{\text{rev}})} = \frac{e^{\beta W}\sum_{i\in\text{supp}(\rho_s)} e^{-E_i}}{\sum_j e^{-E_j}}, \tag{A11}$$

where $\text{supp}(\rho_s)$ is the support of the state ρ_s. Noticing that

$$-D_0(\rho_s||\tau_s) = \ln\sum_{i\in\text{supp}(\rho_s)} e^{-\beta E_i} - \ln\sum_{j\in\text{supp}(\tau_s)} e^{-\beta E_j}, \tag{A12}$$

$$\frac{P(W,\mathcal{P})}{P(-W,\mathcal{P}^{\text{rev}})} = \exp\{\beta W - D_0(\rho_s||\tau_s)\}. \tag{A13}$$

■

Notice that each of the probabilities in the numerator and the denominator of the fraction on the left-hand side (LHS) of Eq. (A13) equal one, i.e.,

$$\begin{aligned}P(W,\mathcal{P}) &= \sum_{i,j} k_{i\to j}\frac{P(E_i)}{g(E-E_i)}\\ &= \sum_i g(E-E_i)\frac{P(E_i)}{g(E-E_i)}\\ &= 1, \end{aligned}$$

and

$$\begin{aligned}P(-W,\mathcal{P}^{\text{rev}}) &= \sum_{i,j} k^{\text{rev}}_{j\to i}\frac{P(E_j)}{g(E-E_j-W)}\\ &= \sum_j g(E-E_j-W)\frac{P(E_j)}{g(E-E_j-W)}\\ &= 1, \end{aligned}$$

since the sum of probabilities of the initial and final energy levels have to equal one. Therefore, the fully deterministic work extraction has only a trivial distribution of one value of extracted work and zero everywhere else. For a fluctuation theorem to say something about the work probability, it has to have a width. Below, we extend the relation to the ϵ-deterministic case, thereby giving the proof of Theorem 1. To achieve this, we use the technique of β ordering as introduced in Supplemental Note 4 of Ref. [3] and as illustrated in Fig. 2. The graph plots the Gibbs rescaled probabilities $P(E_i)e^{\beta E_i}$ against the truncated partition function $\sum_i e^{-\beta E_i}$ for each energy level of the work system E_i, arranging states in decreasing order of their value $P(E_i)e^{\beta E_i}$. In this staircase function, the area of each rectangle is equal to the probability of the state $P(E_i)$ and the total width is equal to Z, the partition function. Removing the maximum number of bins in Fig. 2, the minimum total probability then corresponds to removing states from the right side of the β-ordered spectrum. This procedure is called *smoothing*.

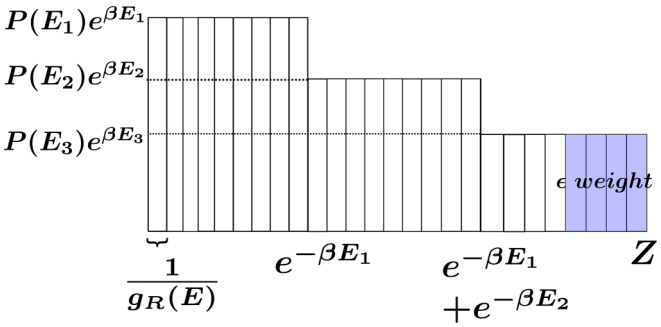

FIG. 2. A weight ϵ is being taken out after β ordering of the initial state. β ordering refers to rearranging the eigenstates' energies in decreasing order of their weight $P(E_i)e^{\beta E_i}$. Since we are interested in removing as many eigenstates as possible while keeping a total weight of at least $1-\epsilon$, we cut from the far right end of the β-ordered spectrum. The white area shows the state that is ϵ-close to the initial state and in our case maximizes the Renyi relative entropy of order 0 in Eq. (1).

Proof. We want to express the LHS of Eq. (3) in terms of the work extracted from the system. We first start by giving the ratio of the work cost probability to the probability of its reverse process for the case where maximum possible work was extracted. Then, we use that to prove the fluctuation theorem in the general case.

Starting with extracting the maximum work in the ϵ-deterministic forward process, we need to find a minimal set of eigenstates such that their total probability is at least $1-\epsilon$. Removing an ϵ weight of the initial state amounts to such a smoothing.

To achieve this, we β-order the eigenstates of the initial state, as explained in the main text. Then, we remove as many eigenstates from the low weight end of the spectrum as possible while staying within the $1-\epsilon$ limit. This is done by choosing an index l, such that

$$1-\epsilon \geqslant \sum_j\sum_{i=1}^{l} k_{i\to j}\frac{P(E_i)}{g(E-E_i)} \tag{A14}$$

and

$$1-\epsilon \leqslant \sum_j\sum_{i=1}^{l+1} k_{i\to j}\frac{P(E_i)}{g(E-E_i)}. \tag{A15}$$

Now, we map these eigenvalues of weight $1-\epsilon$ to W. For such a mapping, the ratio of the work gain distribution to its work cost counterpart is

$$\frac{P(W_{\max},\mathcal{P}^{\epsilon})}{P(-W_{\max},\mathcal{P}^{\epsilon,\mathrm{rev}})} = \frac{\sum_j \left(\left[\sum_{i=1}^{l} k_{i\to j}\frac{P(E_i)}{g(E-E_i)}\right] + k_{l+1\to j}\frac{P(E_{l+1})}{g(E-E_{l+1})}\right)}{\sum_{i,j} k^{\mathrm{rev}}_{j\to i}\frac{P(E_j)}{g(E-E_j-W)}}$$

$$= \frac{\sum_{i=1}^{l} g(E-E_i)\frac{P(E_i)}{g(E-E_i)} + \frac{1-\epsilon-\sum_{i=1}^{l}P(E_i)}{P(E_{l+1})}g(E-E_{l+1})\frac{P(E_{l+1})}{g(E-E_{l+1})}}{\sum_{i,j} k^{\mathrm{rev}}_{j\to i}\frac{P(E_j)}{g(E-E_j-W)}} \tag{A16}$$

$$= \frac{\sum_{i=1}^{l} g(E-E_i)\frac{\sum_j k^{\mathrm{rev}}_{j\to i}\frac{P(E_j)}{g(E-E_j-W)}}{g(E-E_i)} + \frac{1-\epsilon-\sum_{i=1}^{l}P(E_i)}{P(E_{l+1})}g(E-E_{l+1})\frac{\sum_j k^{\mathrm{rev}}_{j\to l+1}\frac{P(E_j)}{g(E-E_j-W)}}{g(E-E_{l+1})}}{\sum_{i,j} k^{\mathrm{rev}}_{j\to i}\frac{P(E_j)}{g(E-E_j-W)}} \tag{A17}$$

$$= \frac{\sum_{i=1}^{l} g(E-E_i) + \frac{1-\epsilon-\sum_{i=1}^{l}P(E_i)}{P(E_{l+1})}g(E-E_{l+1})}{\sum_j g(E-E_j-W^{\epsilon})}. \tag{A18}$$

In Eq. (A16) we used Eqs. (A14) and (A15) to count the number of states with their specific weights that are mapped to the work extracted. As opposed to the setting in Lemma 1, here the initial state of the reverse process and the final state of the forward process are not the same. In Ref. [2], it was discussed that the final state of the system in an ϵ-deterministic process may not be exactly a thermal state. To formulate the fluctuation theorem for these processes, we choose the initial state of the backward process to be the same as that in Lemma 1, where the backward process starts by the system correlated with the bath, such that the reduced state describing the system is in the Gibb state. Using this, in Eq. (A17), the probability of the system being in energy state E_{l+1} is rewritten in terms of the probabilities in the backward process. In the backward process the work W is returned from the battery to the system also via a ϵ-deterministic process. It was shown in Ref. [4] that for ϵ-deterministic processes, the thermal distribution has a maximum work content of $-\frac{1}{\beta}\ln(1-\epsilon)$. This means our reverse process costs this amount less than a deterministic reversal. Therefore, we subtract this amount from the work stored in the battery when calculating the total number of states in the backward process in order to conserve the total energy. Hence, defining $W^{\epsilon} = W + \frac{1}{\beta}\ln(1-\epsilon)$, we have

$$\sum_i k^{\mathrm{rev}}_{j\to i} = g(E-E_j-W^{\epsilon}), \tag{A19}$$

which gives the denominator in Eq. (A18). As we see, Eq. (A18) gives the ratio of the probability of maximum work extraction to that of its reversed process in terms of the number of the eigenstates that are mapped to the work content. However, in an ϵ-deterministic scenario we do not always extract the maximum work possible, which is the motivation for having a fluctuation theorem. We would like to have the ratio above for any possible work that can be extracted in an ϵ-deterministic process. Let us consider an unknown parameter δ to characterize a possible work extracted. A work W^{δ} is obtained by mapping a $1-\delta$ number of eigenstates in Eq. (A18) to work. Since the battery in that case is charged by W^{δ} amount, the ratio above generalizes to

$$\frac{P(W^{\delta},\mathcal{P}^{\epsilon})}{P(-W^{\delta},\mathcal{P}^{\epsilon,\mathrm{rev}})} = \frac{\left(\sum_{i=1}^{l} g(E-E_i) + \frac{1-\epsilon-\sum_{i=1}^{l}P(E_i)}{P(E_{l+1})}g(E-E_{l+1})\right)^{\frac{1-\delta}{1-\epsilon}}}{\sum_j g(E-E_j-W^{\epsilon})}$$

$$= \exp\{\beta W^{\delta} + \ln(1-\delta) - D_0^{\epsilon}(\rho_s||\tau_s)\}. \tag{A20}$$

Equation (A20) follows from a similar argument where we derived Eq. (A13) from Eq. (A11), albeit with smoothing. For the definition of smooth Renyi relative entropy, see Eq. (1). Smoothing maximizes Renyi relative entropy over all states that are ϵ-close to the initial state. As we discussed, in our situation, this is done by removing the eigenstates that have the lowest probability, with the total weight of ϵ. The expression in the large parentheses in the numerator of Eq. (A18) corresponds to $g(E)$ times that entropy and is the same expression used in Ref. [3]. ■

Remark 2. For clarity of the proof we have used the notation W^{δ} for the work extracted. This corresponds to W in the main text and in particular in the statement of Theorem 1.

Remark 3. To formalize the fact that the backward process is the reversal of the forward process in the presence of the epsilon error, notice that we require that the battery has to be charged the amount required to perform the backward process. Although here the global unitary that generates the thermal operation is not unique, one can still decompose any such unitary U as a completely positive trace-preserving map on the reservoir and the system $\mathcal{N}(\cdot)$ and the corresponding translation unitary map $\Gamma(\cdot)$ on the battery as $U\rho_{Rsb}U^{\dagger} = \mathcal{N}_i(\rho_{Rs})\otimes\Gamma_i(\rho_b)$. This means that defining the backward process by the inverse of the forward unitary will send the battery from its charged level to the ground state. Finally, in order for this process to be a valid backward process, we require that the map $\mathcal{N}$ send the final state of the forward process of Lemma 1 to a state that is close to our initial state. Suppose the unitary that has been implemented is one that takes the states that are on the surface of the ball

of ϵ-close states to ρ_s into τ_s, i.e., the final state achieved in Lemma 1. If the process fails to extract the maximum work, the final state will not exactly be τ_s. Nevertheless, the backward process, defined by the inverse of the unitary, will map the state given by the final state of Lemma 1, to a state that is on the surface of the ball of ϵ-close states to ρ_s. This is of course the farthest that the final state of the backward process can be from ρ_s. Suppose the unitary is implemented, such that mapping does not use all the eigenstates in the range depicted as ϵ in Fig. 2. In that case, the inverse process will map the final state of Lemma 1 into a state within the ball of ϵ-close states to ρ_s.

[1] L. Del Rio, J. Åberg, R. Renner, O. Dahlsten, and V. Vedral, Nature (London) **474**, 61 (2011).

[2] J. Gemmer and J. Anders, New J. Phys. **17**, 085006 (2015).

[3] M. Horodecki and J. Oppenheim, Nat. Commun. **4**, 2059 (2013).

[4] J. Åberg, Nat. Commun. **4**, 1925 (2013).

[5] M. Campisi, P. Hänggi, and P. Talkner, Rev. Mod. Phys. **83**, 771 (2011).

[6] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).

[7] G. Gour and R. W. Spekkens, New J. Phys. **10**, 033023 (2008).

[8] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, Phys. Rev. Lett. **111**, 250404 (2013).

[9] Á. M. Alhambra, J. Oppenheim, and C. Perry, Phys. Rev. X **6**, 041016 (2016).

[10] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, Science **296**, 1832 (2002).

[11] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, Nature (London) **437**, 231 (2005).

[12] F. Douarche, S. Ciliberto, A. Petrosyan, and I. Rabbiosi, Europhys. Lett. **70**, 593 (2005).

[13] P. Faist, J. Oppenheim, and R. Renner, New J. Phys. **17**, 043003 (2015).

[14] M. Tomamichel, R. Colbeck, and R. Renner, IEEE Trans. Inf. Theory **55**, 5840 (2009).

[15] O. C. O. Dahlsten, M.-S. Choi, D. Braun, A. J. P. Garner, N. Yunger Halpern, and V. Vedral, New J. Phys. **19**, 043013 (2017).